ABSTRACT

Multistage residual vector quantizers (RVQ) with optimal direct sum decoder codebooks have been successfully designed and implemented for data compression. Due to its multistage structure, RVQ has the ability to densely populate the input space with voronoi cell partitions. The same design concept has yielded good results in the application of image-content classification [1]. Furthermore, the multistage RVQ, with stage-wise codebooks, provides an opportunity to perform fine-grained feature attribution for image understanding, in general, and feature foundation data generation for natural and man-made structure recognition, in specific. In [1], the information at the stages of RVQ is heuristically integrated to perform class conditional pattern recognition; hence the process is not robust. Markov random field (MRF) provides a suitable Bayesian framework to integrate the information available at the various stages of RVQ to achieve optimized classification in the maximum a-posteriori sense (MAP).

Index Terms— Residual vector quantization (RVQ), markov random field (MRF), classification, class-conditional feature attribution, maximum a-posteriori sense (MAP).

1. INTRODUCTION

Multistage RVQs with optimal direct sum decoder codebooks have been successfully designed and implemented for data compression. The same design concept has yielded good results in the application of image-content classification and has also provided an effective platform to perform image driven data mining (IDDM) [1]. The number of codevectors in the stage-wise codebooks and the number of stages in the RVQ are free parameters. They can be varied to adjust the reconstruction error. Equivalently, control over these free parameters of the direct sum codebook allows the RVQ to generate a dense covering over the input space. This feature can prove to be very useful in image understanding and generation of feature foundation data in images which can require multi-classification to recognize features belonging to various classes. The number of stages and/or number of codevectors per stage can be varied to facilitate representation of more than one class with a single codebook. So far, only heuristic methods have been used to integrate the information at the various stages of the RVQ [1]. Consequently, the classification process is not robust. Markov random field (MRF) provides a suitable Bayesian framework to structure multistage RVQs to achieve optimal m-ary classification in the maximum a posteriori (MAP) sense, where \( m = \{1,2,3,\ldots,M\} \). In [9] MRF with a scalar codebook RVQ was implemented for classification application. This paper is focused on the use of MRF on a vector codebook RVQ. In section 2 and 3 the RVQ and MRF models are briefly explained, respectively, followed by section 4 in which the use of the MRF model to structure the optimal direct-sum RVQ for designing MRF-RVQ models for the purpose of classification is explained. Section 5 demonstrates the effectiveness of the proposed MRF-RVQ model in classification followed by a conclusion in section 6.

2. RESIDUAL VECTOR QUANTIZATION

Residual vector quantization or quantizer (RVQ), also known as multistage vector quantization or quantizer (MSVQ), have been designed with direct sum codebooks [2], [3], [4] and [5]. Direct sum codebooks are memory efficient. Fig.1 illustrates the construction of an RVQ with \( N \) stages and two code-vectors per stage wise codebook. The stages are numbered in the top-down manner, where the first stage is the top-most layer and the last stage is the bottom-most layer of the RVQ. Further works focused at improving the direct sum codebook design are discussed in [8]. Common to all these design strategies is sub-optimal sequential search encoding, done so to make the RVQ implementation computationally feasible.

In this paper, MRF framework is imposed on the RVQ developed by C. F. Barnes [2].

3. MARKOV RANDOM FIELD

3.1. Bayesian labeling based on MRF

The MRF model used in this paper is based on S. Z. Li [7]. Let \( s = \{1,2,\ldots,m\} \) be a set of discrete sites and \( L^* = \{0,1,\ldots,M\} \) be a set of labels which include \( M \) physical
labels (1,2,…,M) and a virtual NULL label (0). The aim is to assign a label from $L^*$ to each of the sites in $s$ subject to some contextual constraints. Let $f = \{f_1, f_2, \ldots, f_m\}$ be a configuration of an MRF with $f_i \in L^*$ assuming a mapping $f : l \rightarrow L^*$ or a labeling of $l$. Let $W = L^* \times L^* \times \ldots \times L^*$ (m-tme) be the set of all possible configurations.

Given the likelihood function $p(r \mid f)$ and a priori probability $p(f)$, the posterior probability can be computed by using the Bayesian rule $P(f \mid r) \propto p(r \mid f) p(f)$. The bayesian labeling problem is the following: given the observation $r$, find the map configuration $f^*$ from an admissible space $w$, that is,

$$f^* = \arg\max_{f \in W} P(f \mid r)$$ (1)

According to the Hammersley-Clifford theorem of markov-gibbs equivalence [7],[8], the prior probability $p(f)$ obeys a gibbs distribution

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}(U(f))}$$ (2)

Where $Z$ is a normalizing constant, $T$ is a global control parameter called the temperature and $U(f)$ is the prior energy. The prior energy has the form

$$U(f) = \sum_{\epsilon \in \mathcal{E}} V_c(f)$$ (3)

Where $c$ is the set of cliques in a neighborhood system $\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in l\}$ for $l$ in $\mathcal{N}_i$ is the collection of sites neighboring to $i$.

The likelihood $p(r \mid f)$ depends how $r$ is observed. It can usually be represented in an exponential form $p(r \mid f) = Z^{-1} \times e^{-u(r \mid f)}$, where $U(r \mid f)$ is the likelihood energy. Hence the posterior probability is gibbs distribution $p(f \mid r) = Z^{-1} \times e^{-u(r \mid f)}$ with posterior energy

$$U(f \mid r) = U(f) / T + U(r \mid f)$$ (4)

Therefore, given an observation $r$, a labeling $f$ of sites in $l$ is also an MRF on $l$ with respect to $\mathcal{N}$. The map solution is equivalently found by

$$f^* = \arg\min_{f \in W} U(f \mid r)$$ (5)

### 3.1.1 Posterior distribution

In all cases, $\mathcal{N}_i$ can be the set of all the other sites $j \neq i$. This is a trivial case for mrf. In contextual matching, it can consist of all other sites which are related to $i$ by the observed relations in $r$. When the scene is very large, $\mathcal{N}_i$ needs to include only those of the other sites which are within a spatial distance from $i$ i.e., $\mathcal{N}_i = \{j \neq i \mid \text{dist} \{\text{feature}_i, \text{feature}_j\} < \alpha, j \in l\}$. The threshold $\alpha$ can be reasonable related to the size of the model object. The set of first order cliques is $c_1 = \{\{i\} \mid i \in l\}$. The set of second order cliques is $c_2 = \{\{i,j\} \mid j \in N_i, i \in l\}$. In this paper, only cliques of up to order two are considered.

The single site potential is defined as

$$V_1 = \begin{cases} r_{i0}^\alpha & \text{if } f_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

where, $v_{10}$ is a constant. This definition implies that a penalty $v_{10}$ is incurred, if $f_i$ is the null label; or otherwise no penalty. The two sites potential is defined as

$$V_2(f_i, f_j) = \begin{cases} v_{20} & \text{if } f_i = 0 \text{ or } f_j = 0 \\ 0 & \text{otherwise} \end{cases}$$

Where $v_{20}$ is a constant. Similarly, a penalty $v_{20}$ is incurred if either $f_i$ or $f_j$ is the null; or otherwise no penalty. The above clique potentials specify the prior energy.

Likelihood energy: the joint likelihood function $p(r \mid f)$ has the following characteristics: (i) it is conditioned on pure non-null matches $f_i \neq 0$, (ii) it is independent of the neighborhood system $\mathcal{N}$, and (iii) it depends on how the model object is observed in the scene which in turn depends on the underlying transformation and noise. Assume $r = (r_1, r_2)$ where $r_1 = \{R_1(i) \mid i \in l\}$ and each $r_1(i)$ is a vector of $k_1$ unary properties; $r_2 = \{R_2(i,j) \mid i \in l, i \neq j\}$ and each $r_2(i,j)$ is a vector of $k_2$ binary relations. The same assumptions are also made for $r = (r_1, r_2)$. Assuming an observation model to be $r = r + n$ where $n$ is the independent Gaussian noise, then the likelihood energy is

$$U(r \mid f) = \sum_{i \in l \mid f_i = 0} V_1(r_i \mid f_i) + \sum_{i,j \in l \mid f_i \neq 0} V_2(r_i \mid f_i, f_j)$$ (6)

Because the noise white, we have $U(r \mid f) = U(r \mid f) + U(r \mid f_i, f_j)$ and $U(r \mid f_i, f_j) = U(r_2 : f_i, f_j)$. The likelihood potentials are

$$V_1(r_1(i) \mid f_i) = \sum_{k=1}^{K_1} [r_{1,k}(i) - R_{1,k}(f_i)]^2 / 2\sigma_{1,k}^2$$ (7)

And

$$V_2(r_2(i,j) \mid f_i, f_j) = \sum_{k=1}^{K_2} [r_{2,k}(i,j) - R_{2,k}(f_i, f_j)]^2 / 2\sigma_{2,k}^2$$

where $\sigma_{n,k}^2$ (k = 1,2,⋯,K and n = 1,2) are the standard deviations of the noise components. The vectors $R_1(f_i)$ and $R_2(f_i, f_j)$ are the “mean vector” for the random vectors $r_1(i)$ and $r_2(i,j)$, respectively. When the noise is correlated, there are correlation terms in the likelihood potentials. The assumption of the independent Gaussian noise made may not be accurate but is usually a practical approximation.

The posterior energy in (4) can then be derived as

$$U(f \mid l) = \sum_{(i) \in l \mid f_i = 0} V_1(f_i) / T + \sum_{(i,j) \in l \mid f_i \neq 0 \text{ or } f_j = 0} V_2(f_i, f_j) / T + \sum_{(i) \in l \mid f_i = 0} V_1(r_1(i) \mid f_i) + \sum_{(i,j) \in l} V_2(f_i, f_j)$$
\[ \sum_{(l,j) \in \mathcal{E} \ni f_l \neq 0 \lor f_j \neq 0} V_2 \left( r_2(i,j) \mid f_i, f_j \right) \]  

(8)

It can be written into a compact form

\[ U(f) = E(f) = \sum_{i \in \mathcal{I}} E_1(f_i) + \sum_{i \in \mathcal{I}, \sum_j E_2(f_i, f_j) \right) \]  

(9)

where

\[ E_1(f_i) = \left\{ \begin{array}{ll}
\frac{v_{n0}}{V_1(r_1(i))} & \text{if } f_{i=0} \\
0 & \text{otherwise}
\end{array} \right. \]

and

\[ E_2(f_i, f_j) = \left\{ \begin{array}{ll}
\frac{v_{n0}}{V_2(r_2(lj))} & \text{if } f_{i \neq 0 \lor f_j \neq 0} \\
0 & \text{otherwise}
\end{array} \right. \]

are local posterior energy of order one and two, respectively.

The involved parameters are the noise variances \( \sigma_{n,k}^2 \) and the prior penalties \( v_{n0} \), with \( T = 1 \) fixed. Only the relative, not absolute, values of \( \sigma_{n,k}^2 \) and \( v_{n0} \) are important because the solution \( f^\ast \) remains the same after the energy \( E \) is multiplied by a factor. The \( v_{n0} \) in the MRF prior potential functions, controls the behaviour of the system and can be set as desired. The higher the prior penalties \( v_{n0} \), the fewer features in the scene will be matched to the NULL for the minimal energy solution.

4. MRF STRUCTURED RVQ CLASSIFIER

To implement the MRF, in section 5, on RVQ (referred to as MRF-RVQ), the associated clique structure, clique parameters and the neighborhood system need to be defined. In the MRF-RVQ structure, the MRF is implemented as a relational graph (RG) on the stages of the RVQ. The RG has the stagewise codebooks or the constituent stagewise codevectors as nodes. The unary property of a particular node is the value of the code-vector under consideration. The relation between any two nodes \((b\text{-}i\text{-}nary \text{relation})\) or \(n\) nodes \((n\text{-}ary \text{relation})\) is the probability of an input to transition form a codevector at ‘s’ stage to the codevector at ‘s’ stage. The transition to a codevector at ‘s’ stage is determined by the minimum distance between the codevector at ‘s-1’ stage and the codevector at ‘s’ stage.

4.1. Neighborhood System

The neighborhood system \( \mathcal{N} \) is defined as the number of stages specified in the MRF-RVQ structure. So, if \( \mathcal{N} = \mathcal{N} \), the neighborhood system \( \mathcal{N} \) is composed of \( N \) stagewise codebooks of the RVQ.

4.2. Cliques

In the proposed MRF-RVQ, only pairwise cliques are considered. The binary relation between the pairwise cliques is top-down directional i.e., the transitional probability between any two code-vectors \( cv_i \) and \( cv_j \) such that the mean squared error (MSE) distance between the direct sum output of \( cv_i \) code-vector, at stage \( i \), and \( cv_j \) code-vector is minimum among all the codevectors at stage \( j \), where \( i < j \) and \( i \& j \) are indices of the RVQ stages arranged in a top-down manner.

For the purpose of this particular RVQ framework, the parameters \( \sigma_{n,k}^2 \) and the partition function \( Z \) can be ignored. The clique parameters are all equal to 1. Hence, each path between the pair-wise codevectors, so determined as per minimum MSE criterion, is given equal weight. These weights are also associated with the binary relations between the code-vectors in a pair-wise clique. The binary relations are estimated by calculating the frequency of the mapping of training inputs to the individual code-vectors in the stage-wise codebooks. Intuitively, the MRF-RVQ can be considered as a probability transition cycle between the stage-wise codebooks and the binary relation between the pair-wise clique is, then, the transitional probability between the code-vectors in the pair-wise clique. Figure 1 shows the structure of the MRF-RVQ where the RVQ has \( N \) stages with two code-vectors per stage-wise codebook. Every possible binary relation in a pair-wise clique is denoted by the transitional probability \( P_{ij} \) where \( k \) is the index of the stage, \( i=(1,2) \) and \( j=(1,2) \) are the indices of the code-vectors of the stage-wise codebook such that the \( kP_{ij} \) indicates the transitional probability from the \( i \)th code-vector of the \( k \)-1 stagewise codebook to the \( j \)th code-vector of the codebook at the \( k \)th stage of the MRF-RVQ.

5. EXPERIMENTAL RESULTS

Here the results of MRF-RVQ are presented for \( m \)-ary classification application. In the experiment a direct sum codebook RVQ with the following specifications is used:-

a) Number of Stages \( N = 8 \)
b) Number of codevectors per stage \( P = 4 \)
c) Number of classes \( m = 4 \)
The four classes represent the following objects: 
- a) Class 1 = ‘Vehicle’
- b) Class 2 = ‘Building’
- c) Class 3 = ‘Road and Pavement’
- d) Class 4 = ‘Trees and Grass’

6. CONCLUSION

RVQ can be very effectively used for low level object specific feature classification. It is made possible by varying the number of stages and codevectors-per-stage in the RVQ. The future work in this direction entails a study on the effect of increasing the size of the codevectors or templates. It is anticipated that increasing the size of the codevectors will result in the reduction of false detections as more context will be captured by the codevectors. Moreover, a comparison between RVQ and other template-based classification techniques such as principal component analysis and support vector machines will reveal more details on the effectiveness of the RVQ as a classifier.

7. REFERENCES